

PCNA binding proteins in *Drosophila melanogaster*: the analysis of a conserved PCNA binding domain

Emma Warbrick*, Wayne Heatherington¹, David P. Lane and David M. Glover

CRC Laboratories, Medical Sciences Institute, University of Dundee, Dundee DD1 4HN, UK and

¹SmithKline Beecham Pharmaceuticals, Mundells, Welwyn Garden City, Herts AL7 1EY, UK

Received June 1, 1998; Revised and Accepted July 6, 1998

ABSTRACT

The eukaryotic polymerase processivity factor, PCNA, interacts with cell cycle regulatory proteins such as p21^{WAF1/Cip1} and Gadd45, as well as with proteins involved in the mechanics of DNA repair and replication. A conserved PCNA-binding motif is found in a subset of PCNA-interacting proteins, including p21, suggesting that the regulation of these interactions is important for the co-ordination of DNA replication and repair. We have identified several classes of protein which bind to *Drosophila* PCNA. Two of these proteins contain the consensus PCNA-binding domain: one is the Dacapo protein, a *Drosophila* homologue of p21^{WAF1/Cip1}, and the second is the transposase encoded by the *Pogo* DNA transposon. A conserved PCNA-binding domain is also present in a human relative of *Pogo*, named *Tigger*, suggesting that this domain has a functional role in this class of transposable element. This raises interesting possibilities for a novel method of transposition in which the transposase might be targeted to replicating DNA. Finally, we have investigated the use of this conserved PCNA-binding domain as a predictor of PCNA-binding capacity.

INTRODUCTION

Proliferating cell nuclear antigen (PCNA) plays an essential role in both the replication and repair of DNA. It forms a toroidal shaped trimer which interacts non-specifically with DNA by encircling it, forming a sliding clamp which tethers the polymerase complex to the DNA duplex. PCNA is highly conserved and has been identified in a very wide range of eukaryotes (1,2). Genes showing sequence homology with PCNA have been identified as far down the evolutionary ladder as the Archaea, whose DNA replication and repair machinery appears to be more typically eukaryotic than prokaryotic (3). The β subunit of *Escherichia coli* polymerase III is a functional prokaryotic homologue of PCNA and although this protein shows six-fold symmetry like PCNA, it exists as a dimer rather than a trimer (4).

PCNA is an essential component of the DNA replication machinery, acting as the processivity factor for polymerases δ and ϵ (5,6). In addition to its role in replication, PCNA is also required

for nucleotide excision repair (7,8) and plays a role in one pathway of base excision repair (9). Recent analysis of proteins which interact with PCNA have shown that it interacts not only with enzymes involved in the mechanics of DNA repair and replication, but also binds to cell cycle regulatory proteins such as p21 and Gadd45 (10–14). This suggests that the interaction of PCNA with other proteins is a key regulatory target for the co-ordination of DNA replication and repair.

PCNA in *Drosophila* is encoded by the *mus209* gene: mutations in *mus209* result in lethality, although several *mus209* mutants are temperature sensitive and provide unique tools to analyse the functions of PCNA in a developing multicellular organism (15). PCNA is required throughout development and maternally encoded PCNA is essential for embryogenesis (15). Analysis of the 5'-flanking region of *mus209* indicates that its transcription is regulated by homeodomain proteins, suggesting that its expression is coordinated with cell proliferation and differentiation (16). The *mus209* mutants are highly sensitive to a range of DNA damaging agents, supporting the role of PCNA in DNA repair. The suppression of position effect variation seen in these mutants points to a role for PCNA in chromatin assembly (15).

In a screen for proteins which interact with *Drosophila* PCNA, we have identified several interacting proteins. Two of them, which we describe here, have not previously been identified as PCNA-binding proteins. One is the Dacapo protein, which is a cyclin-dependent kinase inhibitor (17,18). The second is the transposase encoded by the *Pogo* transposon, which belongs to the TC1/Mariner superfamily of transposons (19). We find that both proteins contain a consensus PCNA-binding domain shared by other PCNA-binding proteins such as p21, Fen1, XPG and MCMT (12,20–23). We find that a human relative of *Pogo*, named *Tigger*, also potentially encodes a protein containing a conserved PCNA-binding domain (24). This raises interesting possibilities for a novel method of transposition of such elements, possibly by targeting the transposase to replicating DNA.

MATERIALS AND METHODS

Plasmid expression constructs and yeast two-hybrid methods

Manipulations of *E.coli* and DNA were by standard methods (25). Double-stranded plasmid DNA was sequenced using the Sequenase™ protocol (US Biochemical). The plasmids expressing human, *Drosophila* and *Schizosaccharomyces pombe* PCNA

*To whom correspondence should be addressed. Tel: + 44 1382 344806; Fax: +44 1382 224117; Email: e.warbrick@dundee.ac.uk

have been described (12,14). Growth and maintenance of *Saccharomyces cerevisiae* was according to standard methods (26). Transformation was carried out by the method of Gietz *et al.* (27). The *S.cerevisiae* strain Y190 (MATa *leu2-3,112 ura3-52 trp1-901 his3-D200 ade2-101 gal4D gal80D cyh^R URA3::GAL1-lacZ LYS2::GAL1-HIS3*) was used for all two-hybrid analysis, which expresses the reporter genes *lacZ* (*E.coli*) and *HIS3* (*S.cerevisiae*) under the control of the *GAL1* promoter. All two-hybrid screening and analysis was carried out as previously described (12,20). pACT plasmids identified in the two-hybrid screening experiment were tested against pAS plasmids encoding unrelated Gal4 fusion proteins to exclude false positive results: these included pAS-Snf1, pAS-p53, pAS-Cdk2 and pAS-lamin (20).

Peptide analysis

The following 20 amino acid peptides were used (Chiron Mimotopes, Australia). These peptides were linked via residues SGSG at the N-terminus to biotin. p21, KRRQTSMTDFYHSKRRLLIFS; p21-A, KRRATSMDFYHSKRRLLIFS; Dacapo, RKRQPKITEFMKERKRLAQA; Dacapo-A, RKRAPKITEFMKERKRLAQA; Z50796, TKRQKMTDFMAVSRKKNSL; Z50796-A, TKRQKMTDFMAVSRKKNSL; Pogo, KLFNLHINS AVLQKKITDYF; Pogo-A, KLFNLHINS AVLAKKITDYF; Tigger, LMWQTSLLSYFKKLPPQP; Tigger-A, LMWQTSLLSYFKKLPPQP; Consensus, KKRQKRLTDF-FKRKRKLKEA; Consensus-A, KKRAKRLTDFFKRKRKLKEA; Unrelated, PESVELKWSEPNEELIKFM.

ELISA analysis of PCNA binding to peptides

Plastic plates for enzyme-linked immunoabsorption assay (ELISA) were coated with 100 µl 5 µg/ml streptavidin by drying overnight at 37°C. They were washed with phosphate-buffered saline (PBS) containing 0.2% Tween 20 (PBST) and blocked with 5% non-fat milk powder in PBS for at least 2 h at room temperature. Each well was incubated in turn with the following reagents (the wells were extensively washed in PBST between steps): (i) 0.5 µg peptide diluted in 100 µl 0.1% milk-PBS (2 h, room temperature); (ii) up to 50 µg protein in 100 µl 0.1% milk-PBS, either total cell lysates of *E.coli* BL21 expressing human or *S.pombe* PCNA or purified human or *S.pombe* PCNA (1 h, room temperature); (iii) primary anti-PCNA rabbit polyclonal antiserum 3009 or PC10 monoclonal anti-PCNA antibody diluted 1 in 2000 in 2% milk-PBS (1 h, room temperature); (iv) secondary horseradish peroxidase (HRP)-conjugated anti-rabbit or anti-mouse antibodies diluted 1 in 1000 in 2% milk-PBS (1 h, room temperature); (v) 50 µl chromogenic substrate TMB at 100 µg/ml in 0.1 M sodium acetate, pH 6.0, containing a 1 in 1000 diluted solution of H₂O₂ (30%). Once the colour reaction had developed, the reaction was stopped by adding 50 µl 1 M H₂SO₄ and the plate read using a Dynatech 5000 ELISA plate reader at 450 nm.

Competition ELISA

Peptide ELISA experiments were carried out as described above, with competing peptides added to the PCNA solution before addition to the immobilized, biotinylated peptides. A p21-derived peptide (KRRQTSMTDFYHSKRRLLIFS) or an unrelated control peptide (KPVRLPSIQAIKCAP) were dissolved in DMSO at 20 µg/mg and diluted in DMSO to various concentrations. The

DMSO-diluted peptides were then diluted 1 in 200 into a solution of 0.1% milk-PBST containing purified human PCNA at 5 µg/ml, immediately before adding to the peptide-coated wells.

Peptide pull-down experiments

An aliquot of 2.7 µg each peptide was incubated with 10 µl streptavidin-agarose beads (Sigma) in PBS for 1 h at room temperature, then the beads were washed extensively in PBS and recovered each time by centrifugation. A sample of 20 µl *Drosophila* embryonic extract was added to the washed beads at a final concentration of 1 mg/ml and incubated with the beads on ice for 1 h. The beads were extensively washed in PBS containing 0.05% Tween 20 and bound proteins removed by boiling in SDS loading buffer. Proteins were separated by 15% SDS-PAGE then electrophoretically transferred to PVDF membrane (Amersham). The membranes were blocked in PBS containing 2% skimmed milk for 30 min, then incubated for 1 h with a polyclonal rabbit anti-*Drosophila* PCNA antibody diluted 1 in 1000 in 2% skimmed milk-PBS. After washing, blots were incubated with secondary HRP-conjugated swine anti-rabbit antibodies diluted 1 in 1000 in 2% skimmed milk-PBS for 30 min, followed by washing in PBS containing 0.05% Tween 20. Bound antibody was visualized using the ECL system according to the manufacturer's instructions (Amersham).

Sequence analysis and profile searching

Sequence analysis was carried out using the UWGCG programs at the Daresbury Seqnet facility and BLAST programs at the NCBI web site (27,28). For profile searching, various combinations of the amino acid sequences shown in Figure 7 were aligned using the program PILEUP. These alignments were then used to create a profile with PROFILEMAKE, with the default values of GapWeight = 3.0 and GapLengthWeight = 0.1. The SWISSPROT database Release 35 was searched using PROFILESEARCH with the default values of GapWeight = 4.50 and GapLengthWeight = 0.05; 59999 sequences were examined. Alignments between the profile and the sequences thus found were created with PROFILEGAPS. The various protein homologues shown in Figure 7 were found with the database searching programs BLASTP and FASTA. Protein sequence alignments were created using a combination of PILEUP, LINEUP and BESTFIT and by visual inspection.

RESULTS

Two-hybrid screening identifies PCNA-binding proteins from *Drosophila*

In order to identify *Drosophila* proteins which interacted with PCNA, we employed a yeast-based two-hybrid interaction screening approach. Full-length *Drosophila* PCNA was expressed in yeast as a fusion protein with the DNA-binding domain of *S.cerevisiae* Gal4 (12). This strain was then transformed with a cDNA library in pACT derived from third instar larvae of *Drosophila melanogaster* which expresses hybrid fusion constructs with the transcriptional activation domain of Gal4 (30). Transformants where the two forms of hybrid protein interacted were detected by their ability to form functionally active Gal4 complexes and thus activate Gal4-dependent reporter constructs. Approximately 5 × 10⁶ transformants were screened for their ability to grow on medium containing 3-aminotriazole as a test for

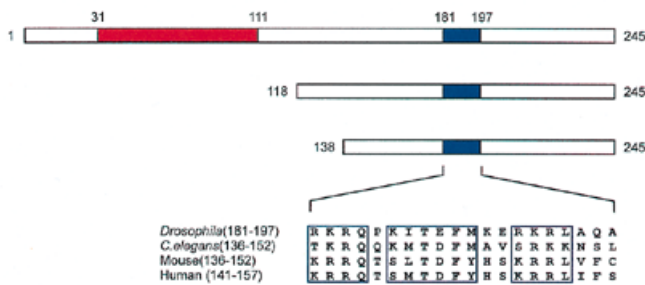


Figure 1. Schematic diagram of the protein sequences encoded by the Dacapo clones identified in the two-hybrid screen against *Drosophila* PCNA. The red box (residues 31–111) represents the consensus cyclin/CDK-binding domain. The blue box indicates the position of the PCNA-binding consensus (residues 118–197) and these residues are shown beneath, along with the conserved domain from *C.elegans*, mouse and human, where the blue boxes indicate highly conserved residues within the domain.

expression of the reporter gene *HIS*³⁺. Of 55 His³⁺ colonies, 20 also expressed detectable levels of β -galactosidase, indicating that the other reporter gene, *lacZ*, was also expressed. The pACT plasmids from these transformants were isolated and tested against plasmids expressing various unrelated protein fusions to test for specificity of interaction (Materials and Methods). From sequence analysis, the clones fell into several classes, two of which are described here.

Pogo and Dacapo proteins bind to the same region of PCNA

Three overlapping clones encoded the Dacapo protein, which has been identified as encoding a cyclin-dependent kinase inhibitor belonging to the p21/p27 class (17,18; Fig. 1). Although, like p27, this protein has a predicted molecular weight of 27 kDa, this identification of Dacapo as a PCNA-interacting protein indicates that it belongs to the same functional family as the p21WAF1/Cip1 subfamily (see Discussion). These three clones contained 1.4, 1.0 and 0.9 kb cDNA inserts, which included coding and 5'-untranslated regions. The smallest Dacapo clone identified in this screen encoded the C-terminal 107 amino acids, which does not include the predicted CDK/cyclin-binding domains. Analysis of this sequence indicated that although the overall homology between p21 family members and Dacapo is not high, specific residues involved in PCNA binding are conserved (Fig. 1; 12,31). Unlike human p21, Dacapo did not interact with human Cdk2 in two-hybrid analysis (data not shown).

Three clones were identified in the two-hybrid screen which contained sequences identical to the *Pogo* transposon (19; Fig. 2). The largest clone starts upstream of the proposed start codon, but is in-frame with the proposed open reading frame (ORF) of the *Pogo* transposase and the second clone encodes amino acids 9–499. The smallest *Pogo* clone encodes only the C-terminal 57 amino acids of the *Pogo* transposase and defines the smallest region necessary for PCNA interaction in this system. Examination of its predicted amino acid sequence identified a region with homology to the consensus PCNA-binding domains described above.

Since PCNA is a highly conserved eukaryotic protein, we used the two-hybrid system to determine whether Dacapo and *Pogo* could bind to PCNA from human and *S.pombe*. We found that in both cases the interaction was evolutionarily conserved (data not

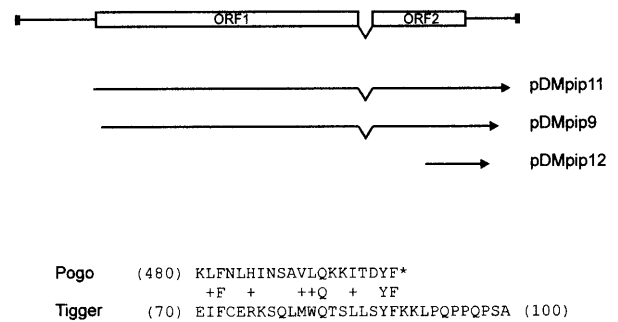


Figure 2. (Top) A schematic diagram of the *Pogo* transposon, showing TIRs as black boxes and the two large ORFs which encode the *Pogo* transposase. The regions corresponding to the inserts identified in the two-hybrid screen against *Drosophila* PCNA are shown. The clones pDMpip11, pDMpip9 and pDMpip12 contained cDNA inserts of ~1.8, 1.6 and 0.3 kb, respectively. **(Bottom)** An alignment between the C-terminal sequences of *Pogo* and the second ORF of *Tigger2*. Identical residues are shown and + indicates a conserved change.

shown). Since these proteins interact with human PCNA, it was possible to test them using the yeast two-hybrid interaction assay against a panel of human PCNA subclones designed for analysing the domain specificity of human PCNA-binding proteins. The results of these experiments show that *Pogo* and Dacapo bind to the same region of PCNA as p21 and Fen1, which includes the interdomain linker region (12,20; Fig. 3).

Pogo and Dacapo proteins contain conserved consensus PCNA-binding domains

Examination of the smallest clones of Dacapo and *Pogo* identified by two-hybrid screening (Figs 1 and 2) reveals that they contain consensus PCNA-binding domains. This is consistent with the observation that they bind to the same region of PCNA as the p21 and Fen1 proteins, which also contain this motif. In order to determine whether these sequences do indeed represent PCNA-binding domains, we obtained 20 amino acid synthetic peptides corresponding to these regions which are linked to biotin through a SGSG linker. The peptides tested represented PCNA-binding domains from p21 homologues in human, *Caenorhabditis elegans* and *Drosophila* (Dacapo) and from the transposases encoded by *Pogo* and *Tigger*, a *Pogo*-related human transposon. A peptide of a completely artificial sequence which represents a consensus PCNA-binding sequence was also tested.

In order to ask whether these peptides were capable of binding to *Drosophila* PCNA, they were conjugated to streptavidin-agarose and incubated with the embryonic extracts. Following extensive washing, bound proteins were analysed by immunoblotting using an anti-*Drosophila* PCNA polyclonal antibody. The results in Figure 4 show that all the peptides examined bound to *Drosophila* PCNA compared with an unrelated peptide, though they showed varying affinities. The peptides derived from human p21, Dacapo, the *Tigger* transposase and the artificially designed consensus peptide all bound to PCNA to a similar degree. The peptide derived from the predicted *C.elegans* p21 homologue (Z50796) showed very poor binding to PCNA following the washing stages, although it did significantly deplete the extract of PCNA in the initial reaction. The *Pogo*-derived peptide, on the

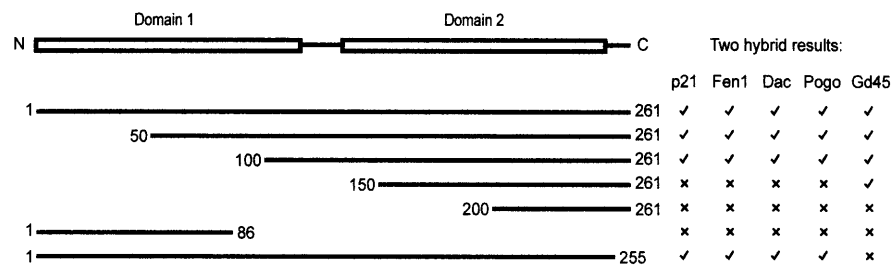


Figure 3. Two-hybrid domain analysis of PCNA binding to Dacapo and Pogo. The PCNA monomer showing the two structural domains is represented schematically. A series of N- and C-terminal deletions of human PCNA (12) were expressed as fusions with the DNA-binding domains of Gal4. These were tested for their interactions with the Dacapo- and Pogo-encoding two-hybrid isolates shown in Figures 1 and 2 as described in Material and Methods: each clone showed the same pattern of interaction. These results were compared with those of previous experiments testing the pattern of interaction of Fen1, p21 and Gadd45 with these PCNA constructs (12,14,20). These results suggest that p21, Fen1, Dacapo and Pogo interact with the same region of PCNA, while Gadd45 shows a different pattern of binding.

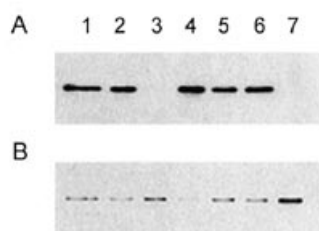


Figure 4. PCNA-binding domain peptides specifically interact with PCNA from *Drosophila* embryonic extracts. Peptides conjugated to streptavidin–agarose beads were incubated with *Drosophila* embryonic extracts and the bound PCNA was analysed by SDS–PAGE followed by western blot analysis with a polyclonal rabbit antibody to *Drosophila* PCNA. The peptides are described in Materials and Methods and the numbered lanes show results as follows; lane 1, p21; lane 2, Dacapo; lane 3, Z50796; lane 4, Pogo; lane 5, Tigger; lane 6, Consensus; lane 7, Unrelated. (A) Western blot results of the PCNA bound to the beads following washing in PBS containing 0.05% Tween 20 and (B) the PCNA remaining in the supernatant following depletion.

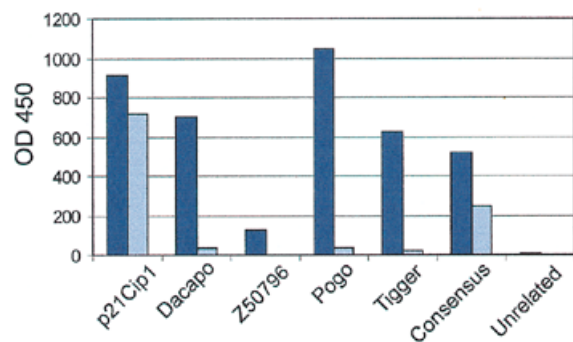


Figure 5. ELISA analysis of peptide binding to PCNA. Biotinylated peptides were immobilized on ELISA plates pre-coated with streptavidin and incubated with purified human PCNA. Binding of PCNA to the immobilized peptides was assessed using the anti-PCNA polyclonal rabbit antiserum 3009, followed by HRP-conjugated anti-rabbit antibody and colorimetric detection at 450 nm. Dark blue bars represent binding to peptides derived from the wild-type protein sequence and light blue bars show results using the equivalent peptide with the conserved Q residue substituted by A.

other hand, showed the highest affinity in this assay, almost completely depleting the extract of PCNA.

In order to investigate the relative affinities of these peptides for PCNA in an *in vitro* assay and to compare them with previously identified PCNA-binding peptides, they were tested for interaction with human PCNA in an ELISA-type assay. In these experiments, peptides were also tested in which the conserved glutamine (Q) was substituted by alanine (A). This residue is not only highly conserved, but in the human p21-derived peptide this mutation abrogates PCNA binding and the inhibition of DNA replication (12). The results in Figure 5 show that each peptide derived from the wild-type protein sequence was capable of binding to human PCNA in this assay, with varying affinities. In particular, the Pogo peptide showed a very high affinity in this assay and the artificially designed peptide also binds strongly to PCNA. These results are consistent with the results shown in Figure 4 for binding to *Drosophila* PCNA. Mutation of the conserved Q residue to A significantly reduced peptide affinity for PCNA in each case, suggesting that this residue is specifically required for the interaction.

In order to examine whether these peptides are binding to the same region of PCNA as human p21, competition ELISA assays were performed. In these experiments, the ability of each peptide to bind to PCNA was tested in the presence of varying

concentrations of the competing, non-biotinylated peptide KRRQTSMTDFYHSKRRLIFS, which is derived from the human p21 protein (Fig. 6). These results indicate that the p21 peptide was able to effectively compete for binding of PCNA to the peptides tested, compared with a control peptide. This strongly suggests that the peptides are binding to the same site within PCNA as p21.

Use of the PCNA consensus as a tool for predicting PCNA-binding proteins

Conserved protein motifs provide useful predictive tools for examining the large numbers of sequences available in the public domain databases. Proteins containing a consensus site can then be tested biochemically for the predicted interaction. In the PCNA-binding motif described here, only a subset of the residues are highly conserved. The BLAST and FASTA algorithms, which are commonly used for database searching, give an identical weighting to each residue of the search sequence, so matched sequences will not necessarily correspond to the consensus. Therefore, to search for best matches to the overall PCNA-binding consensus, we have used a profile search method, in which

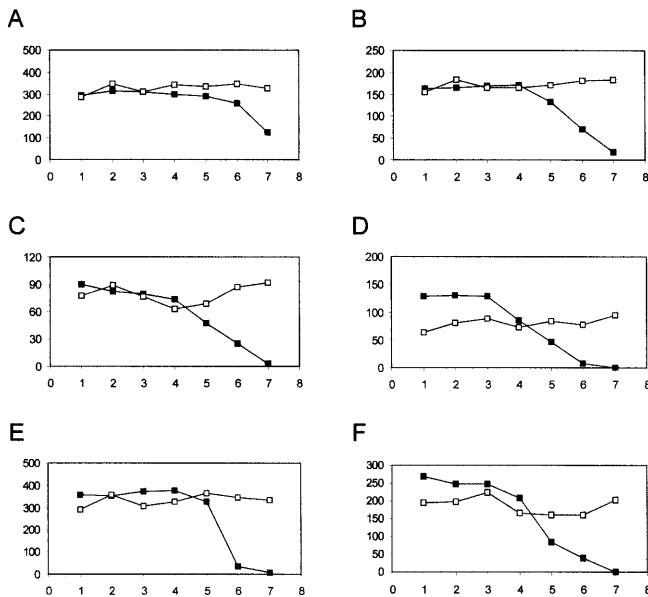


Figure 6. Competition ELISA analysis of peptide-PCNA interactions. Various peptides (A, p21; B, Dacapo; C, Z50796; D, Consensus; E, Pogo; F, Tigger) were immobilized in individual wells and their capacity to bind PCNA in the presence of the non-biotinylated p21-derived peptide KRRQTSMTDFYHSK-RRLIFS was tested by adding this peptide (closed squares) to the PCNA-containing solution immediately before addition to the ELISA plate wells. An unrelated peptide was used as a control for competition (open squares). The OD₄₅₀ (arbitrary units, x-axis), indicating the amount of PCNA bound, is plotted against log₁₀[competing peptide concentration in ng/ml] (y-axis). The p21-derived peptide is able to compete effectively for PCNA binding in each case, compared with the control peptide, strongly suggesting that the p21-derived peptide is binding the same site on PCNA as the immobilized peptide.

sequence alignments are used to construct an amino acid profile, which is then used to search protein sequence databases.

The SWISSPROT database was searched using PROFILE-SEARCH with profiles derived from various combinations of the amino acid sequences shown in Figure 7 (Materials and Methods; 28). A selection of the highest scoring matches in eukaryotic proteins, excluding input sequences, are shown in Figure 8. Although these proteins contain a conserved PCNA-binding domain, it is unlikely that there is a physiological interaction, judging from their function and subcellular localization. We conclude that the presence of a conserved PCNA-binding site does not necessarily indicate a functional interaction.

One notable exception is the identification of a PCNA-binding motif in the uracil-DNA glycosylase (UNG) protein of *S.cerevisiae* (ScUNG), which gave a very high score in several profile searches. This protein functions in base excision repair (BER) to remove mutagenic uracil residues resulting from the deamination of cytosine in DNA (32). Recently, two forms of UNG protein have been identified in human and mouse, which result from alternative splicing events: Ung1 is targeted to mitochondria and Ung2, in which the N-terminal 44 amino acids are encoded by a previously unrecognized exon, is specifically targeted to the nucleus (Fig. 9; 33). There is strong independent evidence that this N-terminal region of Ung2, which contains a conserved consensus PCNA-binding site, plays a role in nuclear targeting,

p21WAF1/Cip1			
Human	P38936	(141)	KRRQTSMTDFYHSKRRRLIFS
Mouse	P39689	(136)	KRRQTSMTDFYHSKRRRLIFS
<i>C.elegans</i>	Z50796	(136)	TKRQKMTDFMAVSRKKNSL
<i>Drosophila</i>	U68477	(181)	RKRQPKITDFMKERKRLAQA
Consensus			RQ TDF R L
Fen1			
Human	P39748	(334)	GSTQGRLLDFFKVTGSLSSA
Mouse	P39749	(334)	GSTQGRLLDFFKVTGSLSSA
<i>Xenopus</i>	U68141	(334)	GSTQGRLLDFFKVTGSLSSA
<i>Drosophila</i>			AQTQVRLDSFFKFLPSTPNA
<i>C.elegans</i>	U13019	(349)	TVAQTRIPDFFFAATSKNFT
<i>S.pombe</i>	P39750	(335)	TIPQGRLLDFFKVPVSPKK
<i>S.cerevisiae</i>	P26793	(334)	SGIQGRLLDFFQVVPKTEQ
<i>M.jannaschii</i>	C64480	(315)	KTQKTLDAWFK*
Consensus			QGRLLD FF
XPG			
Human	P28715	(987)	QQTQLRIDSFRRLAQKEKED
Mouse	P35689	(1000)	HQTQLRIDSFRRLAQKEKQD
<i>Xenopus</i>	P14629	(1020)	QQTQLRIDSFRRLEQHEAAG
<i>C.elegans</i>	AF039713	(726)	GSQMLRLDSFFAQKLDSSLE
<i>S.pombe</i>	P28706	(980)	VGTCGNLTDFEFGNTNVYA
<i>S.cerevisiae</i>	P07276	(992)	KGKQKRLNEFFPREYISGDK
Consensus			QTQ RID FF
MCMT			
Human	P26358	(40)	STRQTITSHFAKGPARRKP
Mouse	P13864	(158)	TTRQTITTAHFTKGPTRRRP
<i>Xenopus</i>	D78638	(149)	AGKQPTILSMFSKSTKRKS
Consensus			Q TI F KG KRK
Transposons			
Pogo	S20478	(489)	AVLQKKITDYF*
Tigger Cds2	U49973	(86)	LMRQTSLLSYFKKLPPQPQ
Consensus			Q YF

Figure 7. A conserved PCNA-binding motif. PCNA-binding domains are shown in protein homologues of p21, Fen1, XPG, MCMT and Pogo transposase. The accession number is indicated in the second column where appropriate and the number in brackets in the third column indicates the position of the first amino acid residue shown (where known). Identical residues are highlighted in blue.

although any interaction with PCNA awaits biochemical analysis. (33,34).

A number of the proteins described here are known by immunofluorescence studies to be targeted (like PCNA itself) to sites of DNA replication and/or repair (35–37; E.Warbrick, P.Coates and P.A.Hall, unpublished observations). We therefore searched for putative PCNA-binding domains in proteins known to be targeted to replication foci. In this way, a consensus PCNA-binding motif was found at the N-terminus of DNA ligase I, which plays an essential role in both DNA replication and nucleotide excision repair and is known to be targeted to replication foci (39,40). The predicted PCNA-binding motif is close to the N-terminus of the protein and is conserved in a wide range of eukaryotes (Fig. 9). Residues 2–115 of human DNA ligase I are essential for replication foci localization and fine mapping has identified a bipartite targeting sequence involving amino acids 1–28 and 111–179 (41,42). More recently, an interaction between the N-terminal 118 amino acids of human DNA ligase I and PCNA has been demonstrated, which can be disrupted by p21 (43). This evidence provides strong support for our hypothesis that DNA ligase I interacts with PCNA through the conserved motif shown

Protein	Accession	Function	Localisation	Organism	Sequence
ScUNG	P12887	Uracil-DNA glycosylase	Nuclear	<i>S.cerevisiae</i>	(18) KRKQTTIEDFFGTTKKSTMEA (37)
Ktr2	P33550	Mannosyltransferase	Membrane	<i>S.cerevisiae</i>	(35) YQRQLSLDSYFFRRSREVSS (54)
Sun2	P40016	Proteasome component	Cytoplasmic	<i>S.cerevisiae</i>	(336) SNMQKSLLPYHLLTKAVKLS (335)
Rnrh	P08056	Ribonuclease RH	Secreted	<i>R.niveus</i>	(180) TEMQSAIESYFPAKAKIDCS (199)
Ssp2	P20613	Sex specific storage protein 2	Secreted	<i>B.mori</i>	(35) VEKQKKILSFFQDVSLNTD (54)
Map3	P31397	Pheromone M-factor receptor	Membrane	<i>S.pombe</i>	(183) WRRQRELQQFFQRDSQLTSK (202)
MI1a	P10147	Macrophage inflammatory protein	Secreted	<i>H.sapiens</i>	(41) QIPQNFADYFETSSQCSKP (60)
UcrQ	P50523	Cytochrome B-C1 reductase subunit VIII	Mitochondrial	<i>S.pombe</i>	(32) PFQQRPMAGFFKTSTQNMFR (51)
Prof	P49230	Profilin	Cytoskeletal	<i>E.histolytica</i>	(94) KTKQAVIIGYFSDASVSAGQ (113)
Shr3	Q02774	Secretory component protein	ER	<i>S.cerevisiae</i>	(161) WDLQKRLKEFYAQEEREAAAN (180)
Sorecin	P30626	Calcium binding	Cytoplasmic	<i>H.sapiens</i>	(29) GQTQDPLYGYYFAAVAGQDGG (48)
Crg4	P55941	Gamma crystallin IV	Eye lens	<i>X.laevis</i>	(94) GDFQGMMEFFDDCPNTYDR (113)
Ncf1	Q09014	Neutrophil NADPH oxidase factor 2	Cytoplasmic	<i>M.musculus</i>	(88) ESRQGTLTLEYFNGLMGLPVK (107)
Dha8	P48448	Aldehyde dehydrogenase 8	Mitochondrial	<i>H.sapiens</i>	(181) PALQSTITRFYGDDEPQSSPN (200)

Figure 8. Results of profile searching for PCNA-binding consensus motifs. Various combinations of the 20 amino acid sequences shown in Figure 7 were used to draw up profiles and these were used to search the SWISSPROT database (Materials and Methods for details). A selection of the best eukaryotic protein matches, excluding input sequences, are shown. All data shown here is described in the database entry for which the accession number is shown. Residues shown in blue are equivalent to those which have direct PCNA interactions in the p21-derived PCNA-binding region and which are highly conserved between the sequences shown in Figure 7. The function and subcellular localization of the majority of these proteins indicate that they are unlikely to interact with PCNA, with the notable exception of *S.cerevisiae* ScUNG (see text for discussion).

Uracil-DNA glycosylase

hUNG1	(1)	MGVFC LGPWGLGRKLRTPGKGFLQLLSRLC---GDHLQA (36)
mUNG1	(1)	MGV-----LGRRLSL--LARRAGLRSLTPNPDSDSRQA (31)
hUNG2	(1)	MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAA (45)
mUNG2	(1)	MIGQKTLYSFFSPTPTGKRTTRSPSEP--VPGSGVAA--EIGGDVA (42)
ScUNG	(1)	MWCMRRLPTNSVMTVARKRKQTTIEDFF---GTTKSTNEAPNKKGKSGATF--MTITNGAA (56)

DNA Ligase I

Human	(1)	MQRSIMSFFHPKKEGKAKKPEKEASN (27)
Mouse	(1)	MQRSIMSFFQPTKEGKAKKPEKETPS (27)
Xenopus	(1)	MQRTIKSFFQPKLGAEVKTKEEKVKN (27)
S.pombe	(1)	MRTVFSQIPRFKQVNQYIRMSTRQSDISNFFISSASHKSEHVEVSQSS (48)
S.cerevisiae	(1)	MRRLLTGCLLSSARPLKSRLPFLMSSSPSSAGKKPKQATLARFFFTSMKNKPTGTPSPKKS (61)
Consensus		MQR I SFF P K K K E S

Figure 9. Predicted PCNA binding motifs. **(Top)** Alignment of the N-terminal regions of uracil-DNA glycosylase (UNG) proteins from human, mouse and *S.cerevisiae* (33,62), showing that a PCNA-binding motif is conserved in the specifically nuclear forms of UNG in mouse and human and in the yeast protein. Ung1 and Ung2 proteins are aligned up to the common splice site (#) and the *S.cerevisiae* sequence is aligned beneath. Identical residues are shown in colour and the stars indicate the conserved residues of the predicted PCNA-binding motif. **(Bottom)** Alignment of the N-terminal regions of DNA ligase I homologues. The proteins shown are (accession no. in brackets) from human (M36067), mouse (P37913), *Xenopus* (P51892), *S.pombe* (P12000) and *S.cerevisiae* (P04819). As above, identical residues are shown in colour and the stars indicate the conserved residues of the predicted PCNA-binding motif.

in Figure 7. While this paper was in preparation, another group also independently identified this motif and showed that the activity of DNA ligase I can be inhibited by its interaction with PCNA (44).

DISCUSSION

Dacapo is a member of the p21WAF1/Cip1 family of CDK inhibitors

The p21/p27 family of CDK inhibitors are multifunctional proteins which contain several functional domains. The N-terminal region of these proteins contains domains that interact with

CDK/cyclin complexes, which are central to the regulation of the cell cycle (45,46). Members of the p21 family are also capable of binding to PCNA through a region in the C-terminus and this domain is capable of inhibiting DNA replication in a PCNA-dependent manner (12,47,48). Although identified as cyclin-dependent kinase inhibitors, members of the p21/p27 family can also promote the assembly of CDKs and cyclins and p21 (at low molar ratios) can stimulate the kinase activity of these complexes (49). Though p21, p27 and p57 are all capable of directing accumulation of Cdk4 and cyclin D in the nucleus, p21 is particularly efficient and the C-terminal region of the molecule is required for this process. This region contains a bipartite nuclear

localization signal which is coincident with the PCNA-binding domain, suggesting that this region may be important somehow for targeting p21 to specific sites within the nucleus. This issue is complicated, however, by the existence of an adjacent, though distinct, cyclin/CDK inhibitory domain in the C-terminus of p21 (50). An attractive model can be proposed in which, rather than representing the sole point of contact with PCNA, the PCNA-binding motif contained within p21, Fen1, XPG, MCMT, etc. is involved primarily in targeting the proteins to the DNA replication machinery within the nucleus.

The Dacapo protein in *D.melanogaster* has been described as belonging to the p21/p27 family of cyclin-dependent kinase inhibitors (17,18). Here we describe identification of the protein by its ability to bind to PCNA, which places it in the same functional class as p21. Dacapo encodes an inhibitor of cyclin E/Cdk2 and is required to regulate cell cycle arrest of epidermal cells at the correct stage of embryogenesis. Unlike p21 in mammalian systems, Dacapo is required for normal embryonic development; loss of function mutants are lethal, with severe developmental defects due to overproliferation (17,18,51). Human p21 and Dacapo do not show a high overall homology (25% identity), rather, the homology is restricted to the regions required for CDK, cyclin and PCNA interactions. Within the PCNA-binding motif, it is predominantly those residues required for the interaction which are conserved (Fig. 1; 12,31).

The Pogo transposase interacts with PCNA

Pogo family transposons were first identified in *Drosophila*, where they are no longer active (19). They are classified as DNA or class II transposons which move by excision and re-integration into the genome, without an RNA intermediate. Such elements are flanked by terminal inverted repeats (TIRs) and encode a transposase protein which binds specifically to the TIRs and catalyses the cutting and pasting events which result in transposition. Mobility is dependent only upon the presence of the TIRs and a catalytically active transposase protein, so non-autonomous elements are as likely to be transposed as autonomous elements.

Recently, a study of the molecular archaeology of human interspersed repeats with TIRs has revealed the presence of several groups with the characteristics of type II DNA-mediated transposons (24). These transposons are no longer mobile and have been described as 'molecular fossils'. The human TIR-flanked repeats were analysed using consensus sequences from multiple alignments and one group showed significant homology to *Pogo*, both in the sequence of the TIRs and the coding sequence (24). This group contained two classes of element: *Tigger1*, so called as they represent a human *Pogo* (52), was primarily represented by full-length elements, of which there are an estimated 3000 copies in the human genome; *Tigger2* is less well characterized and its coding sequence cannot be unambiguously defined, as it is less abundant and predominantly represented by internal deletion products.

Here we describe the identification of a PCNA-binding motif in the *Pogo* transposase and find that this motif is conserved in the second ORF of the *Tigger1* transposon. Equivalent domains from both proteins are functional in binding to both *Drosophila* and human PCNA. This evidence points to a specific function for the domain, which raises many interesting issues concerning the role of PCNA binding in transposase function. Since neither *Tigger* nor *Pogo* transposons appear to be capable of transposition,

functional studies cannot be undertaken. However, it is tempting to speculate that the transposase may utilize the relatively open conformation of DNA around the replication fork to enhance integration and/or excision. Type II DNA transposons move by a cut-and-paste mechanism, so they do not proliferate unless excision is followed by gap repair or the element moves from a replicated to a still unreplicated part of the genome (53,54). This raises the intriguing possibility that the interaction of the *Pogo*/*Tigger* transposases with PCNA may be involved in targeting integration to unreplicated portions of the genome. *Pogo* contains two ORFs which are spliced to give a single ORF encoding a predicted protein of 499 amino acids. The identification of a functional PCNA-binding site in the second ORF of *Tigger1* suggests that, as is the case in *Pogo*, a single *Tigger1* transcript was spliced to encode a single polypeptide.

Putative transposases in other organisms also show homology to *Pogo* and *Tigger*. The most highly related are the Pot2 and Fot1 fungal transposases, which do not appear to contain a consensus PCNA-binding domain, suggesting that an interaction with PCNA is not a feature of all members of this family of transposons (54,55). The *Tigger* and *Pogo* transposases also share homology with cellular proteins, for example Pdc1 in *S.cerevisiae* and its *Kluyveromyces lactis* homologue, Rag3, which are transcription factors, and the murine proteins Jerky and CENP-B, which is a eukaryotic centromere-associated protein (55–58). However, in these cellular proteins, conserved residues known to be involved in the transposase active site are mutated, suggesting that transposase gene-encoded proteins have been adapted to a cellular function. It is possible that the PCNA binding site in *Pogo* and *Tigger* transposases was derived from a cellular PCNA-binding protein through an excision event.

Molecular analysis of the PCNA interaction

The PCNA-binding motifs shown in Figure 7 all share the conserved residues QXX(h)XX(a)(a), where X is any amino acid, (h) indicates residues with moderately hydrophobic side chains, such as leucine, isoleucine or methionine, and (a) indicates residues with highly hydrophobic aromatic side chains, e.g. phenylalanine and tyrosine (60). p21 peptide-PCNA co-crystallization analysis shows that these conserved residues are involved in direct interactions with the PCNA molecule: within the peptide ¹³⁹GRKRRQTSMTDFYHSKRRLIFS¹⁶⁰ the polar side chains of Q144 and T145 form hydrogen bonds with the main chain carbonyls of A252 and P253 in PCNA. The residues ¹⁴⁶SMTDFY¹⁵¹ adopt a helical conformation, with M147, F150 and Y151 interacting with a hydrophobic cavity formed under the connector loop of PCNA (31).

In the p21 peptide, the residues ¹³⁹GRKRR¹⁴³ are involved in poorly ordered interactions with the acidic C-terminus of PCNA (31). Human and mouse DNA ligase I homologues do not have any amino acids N-terminal of the conserved Q, suggesting that in these cases, there may not be an interaction with the C-terminus of PCNA. The C-terminal amino acids of the p21 peptide ¹⁵²HSKRRLIFS¹⁶⁰ form an antiparallel β -sheet with the connector loop and these amino acids have been shown to be important in the inhibition of DNA replication by this peptide (12,31). In the *Pogo*-derived peptide, which has a high affinity for PCNA, these residues are absent. This indicates that the β -sheet interaction with the interdomain linker region on PCNA is not essential. However, this linker domain is also involved in forming

the hydrophobic pocket with which the X(h)XX(a)(a) motif interacts, so is still essential for the interaction, as confirmed by the results in Figure 3.

Summary

We present evidence that Dacapo and the transposase proteins encoded by *Pogo* and *Tigger* bind to the same site on PCNA as do p21 and Fen1, through a conserved QXX(h)XX(a)(a) motif. Further, we show that an entirely synthetic peptide can be designed based on consensus protein sequences which has a high affinity of binding to PCNA. However, a close examination of the PCNA-binding sequences suggests that the molecular interactions of the proteins described here are varied outside the core PCNA-binding domain. An analysis of the exact mechanisms of how these proteins interact with PCNA and how the interactions are coordinated and regulated will be important in understanding the role PCNA plays as a target for cell cycle regulation.

ACKNOWLEDGEMENTS

We thank Steve Elledge for the two-hybrid constructs and library, Daryl Henderson for *Drosophila* PCNA cDNA and for anti-*Drosophila* PCNA antibodies, Nikolai Zhelev for purified human PCNA, Kathryn Ball for peptides, Nicole Gnadt for communicating results prior to publication and Álvaro Tavares for *Drosophila* embryonic extracts and helpful advice. This work was supported by grants from the Cancer Research Campaign and the Association for International Cancer Research.

REFERENCES

- Kelman, Z. (1997) *Oncogene*, **14**, 629–640.
- Jonsson, Z.O. and Hubscher, U. (1997) *BioEssays*, **19**, 967–975.
- Olsen, G.J. and Woese, C.R. (1997) *Cell*, **89**, 991–994.
- Kong, X.-P., Onrust, R., O'Donnell, M. and Kuriyan, J. (1992) *Cell*, **69**, 425–437.
- Hindges, R. and Hubscher, U. (1997) *Biol. Chem.*, **378**, 345–362.
- Lee, S.H., Pan, Z.Q., Kwong, A.D., Burgers, P.M. and Hurwitz, J. (1991) *J. Biol. Chem.*, **266**, 22707–22717.
- Shivji, M.K.K., Kenny, M. and Wood, R.D. (1992) *Cell*, **69**, 367–374.
- Nichols, A.F. and Sancar, A. (1992) *Nucleic Acids Res.*, **20**, 2441–2446.
- Matsumoto, Y., Kim, K. and Bogenhagen, D.F. (1994) *Mol. Cell. Biol.*, **14**, 6187–6197.
- Waga, S., Hannon, G.J., Beach, D. and Stillman, B. (1994) *Nature*, **369**, 574–578.
- Flores-Rozas, H., Kelman, Z., Dean, F.B., Pan, Z.-Q., Harper, J.W., Elledge, S.J., O'Donnell, M. and Hurwitz, J. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 8655–8659.
- Warbrick, E., Lane, D.P., Glover, D.M. and Cox, L.S. (1995) *Curr. Biol.*, **5**, 275–282.
- Smith, M., Chen, I., Zhan, Q., Bae, I., Chen, C., Gilmer, T., Kastan, M., O'Connor, P. and Fornace, A. (1994) *Science*, **266**, 1372–1379.
- Hall, P.A., Kearsley, J.M., Coates, P.J., Norman, D.G., Warbrick, E. and Cox, L.S. (1995) *Oncogene*, **10**, 2427–2433.
- Henderson, D.S., Banga, S.S., Grigliatti, T.A. and Boyd, J.B. (1994) *EMBO J.*, **13**, 1450–1459.
- Yamaguchi, M., Nishida, Y., Moriuchi, T., Hirose, F., Hui, C.-C., Suzuki, Y. and Matsukage, A. (1990) *Mol. Cell. Biol.*, **10**, 872–879.
- Lane, E.M., Sauer, K., Wallace, K., Jan, Y.N., Lehner, C.F. and Vaessin, H. (1996) *Cell*, **87**, 1225–1235.
- de Nuij, J.C., Letendre, M.A. and Hariharan, I.K. (1997) *Cell*, **87**, 1237–1247.
- Tudor, M., Lobočka, M., Goodell, M., Pettitt, J. and O'Hare, K. (1992) *Mol. Gen. Genet.*, **232**, 126–134.
- Warbrick, E., Lane, D.P., Glover, D.M. and Cox, L.S. (1997) *Oncogene*, **14**, 2313–2321.
- Li, X., Li, J., Harrington, J., Lieber, M.R. and Burgers, P.M.J. (1995) *J. Biol. Chem.*, **270**, 22109–22112.
- Gary, R., Ludwig, D.L., Cornelius, H.L., MacInnes, M.A. and Park, M.S. (1997) *J. Biol. Chem.*, **272**, 24522–24529.
- Chuang, L.S., Ian, H., Koh, T., Ng, H., Xu, G. and Li, B.L.F. (1997) *Science*, **277**, 1996–2000.
- Smit, A.F.A. and Riggs, A.D. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 1443–1448.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Rose, M.D., Winston, F. and Hieter, P. (1990) *Methods in Yeast Genetics: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Gietz, D., Stjean, A., Woods, R.A. and Schiestl, R.H. (1992) *Nucleic Acids Res.*, **20**, 1425.
- Genetics Computer Group (1994) *Program Manual for the Wisconsin Package, Version 8*. GCG, Madison, WI.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
- Durfee, T., Becherer, K., Chen, P.-L., Yeh, S.-H., Yang, Y., Kilburn, A.E., Lee, W.-H. and Elledge, S.J. (1993) *Genes Dev.*, **7**, 555–569.
- Gulbis, J.M., Kelman, Z., Hurwitz, J., O'Donnell, M. and Kuriyan, J. (1996) *Cell*, **87**, 1–20.
- Lindahl, T. (1993) *Nature*, **362**, 709–715.
- Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhaus, T.A., Skorpen, F. and Krokan, H.E. (1997) *Nucleic Acids Res.*, **25**, 750–755.
- Slupphaug, G., Markussen, F.-H., Olsen, L.C., Aasland, R., Aarsaether, N., Bakke, O., Krokan, H.E. and Helland, D. (1993) *Nucleic Acids Res.*, **19**, 5131–5137.
- Bravo, R. and Macdonald-Bravo, H. (1987) *J. Cell Biol.*, **105**, 1549–1554.
- Li, R., Hannon, G.J., Beach, D. and Stillman, B. (1996) *Curr. Biol.*, **6**, 189–199.
- Leonhardt, H., Page, A.W., Weier, H. and Bestor, T.H. (1994) *Cell*, **71**, 865–873.
- Lasko, D.D., Tomkinson, A.E. and Lindahl, T. (1990) *J. Biol. Chem.*, **265**, 12618–12622.
- Tomkinson, A.E. and Levin, D.S. (1997) *BioEssays*, **19**, 893–901.
- Montecucco, A., Savini, E., Weighardt, F., Rossi, R., Ciarrocchi, G., Villa, A. and Biamonti, G. (1995) *EMBO J.*, **14**, 5379–5386.
- Cardoso, M.C., Joseph, C., Rahn, H.-P., Reusch, R., Nadal-Ginard, B. and Leonhardt, H. (1997) *J. Cell. Biol.*, **139**, 579–587.
- Levin, D.S., Bai, W., Yao, N., O'Donnell, M. and Tomkinson, A.E. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 12863–12868.
- Jonsson, Z.O., Hindges, R. and Hubscher, U. (1998) *EMBO J.*, **17**, 2412–2425.
- Xiong, Y., Zhang, H. and Beach, D. (1992) *Cell*, **71**, 504–514.
- Harper, J.W., Elledge, S.J., Keyomarsi, K., Dynlacht, B., Tsai, L., Zhang, P., Dobrowolski, S., Bai, C., Connell-Crowley, L., Swindell, E., Fox, M.P. and Wei, N. (1995) *Mol. Biol. Cell.*, **6**, 387–400.
- Goubin, F. and Ducommun, B. (1995) *Oncogene*, **10**, 2281–2287.
- LaBauer, J., Garrett, M.D., Stevenson, L.F., Slingerland, J.M., Sandhu, C., Chou, H.S., Fattaey, A. and Harlow, E. (1997) *Genes Dev.*, **11**, 847–862.
- Chen, J., Chen, S., Saha, P. and Dutta, A. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 11597–11602.
- Ball, K.L., Lain, S., Fahraeus, R., Smythe, C. and Lane, D.P. (1996) *Curr. Biol.*, **7**, 71–80.
- Brugarolas, J., Chandrasekaran, C., Gordon, J.I., Beach, D., Jacks, T. and Hannon, G.J. (1995) *Nature*, **377**, 552–557.
- Milne, A.A. (1928) *The House at Pooh Corner*. Sutton, New York, NY.
- Plasterk, R.H.A. (1995) In Sherratt, D.J. (ed.), *Mobile Genetic Elements*. Oxford University Press, Oxford, UK.
- van Luenen, H.G.A.M., Colloms, S.D. and Plasterk, R.H.A. (1994) *Cell*, **79**, 293–301.
- Kachroo, P., Leong, S.A. and Chattoo, B.B. (1994) *Mol. Gen. Genet.*, **245**, 339–348.
- Daboussi, M.J., Langin, T. and Brygoo, Y. (1992) *Mol. Gen. Genet.*, **232**, 12–16.
- Hohman, S. (1993) *Mol. Gen. Genet.*, **241**, 657–666.
- Prior, C., Tizzani, L., Fukuhara, H. and Wesolowski-Louvel, M. (1996) *Mol. Microbiol.*, **20**, 765–772.
- Toth, M., Grimsby, J., Buzaki, G. and Donovan, G.P. (1995) *Nat. Genet.*, **11**, 71–75.
- Earnshaw, W.C., Sullivan, K.F., Machlin, P.S., Cooke, C.A., Kaiser, D.A., Pollard, T.D., Rothfield, N.F. and Cleveland, D.W. (1987) *J. Cell Biol.*, **104**, 817–829.
- Warbrick, E. (1998) *BioEssays*, **20**, 195–199.
- Percival, K.J., Klein, M.B., Burgers, P.M. (1989) *J. Biol. Chem.*, **264**, 2593–2598.